

Diseño e Implementación de un Sistema de Soporte de Decisiones para el Centro de Documentación Regional “Juan Bautista Vázquez”

Valeria Alexandra Haro Valle¹, Wilson Rodrigo Pérez Rocano¹, Lorena Siguenza-Guzman^{1, 2}, Dirk Cattrysse², Víctor Hugo Saquicela Galarza¹

¹ Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva Ciudadela Universitaria, Cuenca, Ecuador, 01.01.168.

² Centre for Industrial Management, Traffic and Infrastructure, University of Leuven, Celestijnenlaan 300A, Leuven, Belgium, BE-3001.

Corresponding author: {valeria.harov,wilson.perezr}@ucuenca.ec, {lorena.siguenzaguzman, dirk.cattrysse}@kuleuven.be, {lorena.siguenza,victor.saquicela}@ucuenca.edu.ec

Fecha de recepción: Septiembre 21, 2014.

Fecha de aceptación: Octubre 17, 2014.

ABSTRACT

In recent years, volume of library-related data has increased tremendously, as well as complexity of data sources and formats has been escalating. This information explosion has created a big challenge for data managing, archiving and accessing, especially in support of library decision-making. Knowing that a good library management involves integrating a number of strategic indicators, the implementation of a Data Warehouse (DW), that properly manages such amount of information and the complex mix of data sources, becomes an interesting alternative to be considered. This article describes the design and implementation of a decision support system based on a DW approach for the Regional Documentation Centre “Juan Bautista Vazquez”. To assure that all relevant data sources are included during the data analysis, the study utilizes a holistic methodology, previously proposed by Siguenza-Guzman et al. for an integrated library evaluation. This methodology assesses the library collection and services by incorporating important elements for library management, such as service performance analysis, service quality control, collection usage analysis, and information retrieval quality. Based on this data analysis, the study proposes a DW architecture to integrate, process and store the relevant data. These stored data are finally analyzed and visualized by the so-called online analytical processing (OLAP) tools. Initial tests of the implemented decision support system confirm the feasibility and effectiveness of the DW based approach, by successfully integrating multiple and heterogeneous sources, formats and varieties of information, permitting library managers to generate personalized reports, and even allowing to debug the day-to-day transactional processes.

Keywords: Data Warehouse, Decision Support System, OLAP, Hefesto, Pentaho.

RESUMEN

El volumen de datos en bibliotecas ha aumentado enormemente en los últimos años, así como también la complejidad de sus fuentes y formatos de información, dificultando su gestión y acceso, especialmente como apoyo en la toma de decisiones. Sabiendo que una buena gestión de bibliotecas involucra la integración de indicadores estratégicos, la implementación de un Data Warehouse (DW), que gestione adecuadamente tal cantidad de información, así como su compleja mezcla de fuentes de datos, se convierte en una alternativa interesante a considerar. El artículo describe el diseño e implementación de un sistema de soporte de decisiones (DSS) basado en técnicas de DW para la biblioteca de la Universidad de Cuenca. Para esto, el estudio utiliza una metodología holística,

propuesto por Siguenza-Guzman et al., para la evaluación integral de bibliotecas. Dicha metodología evalúa la colección y los servicios, incorporando importantes elementos para la gestión de bibliotecas, tales como: el desempeño de los servicios, el control de calidad, el uso de la colección y la interacción con el usuario. A partir de este análisis, se propone una arquitectura de DW que integra, procesa y almacena los datos. Finalmente, estos datos almacenados son analizados y visualizados a través de herramientas de procesamiento analítico en línea (OLAP). Las pruebas iniciales de implementación confirman la viabilidad y eficacia del enfoque propuesto, al integrar con éxito múltiples y heterogéneas fuentes y formatos de datos, facilitando que los directores de bibliotecas generen informes personalizados, e incluso permitiendo madurar los procesos transaccionales que diariamente se llevan a cabo.

Palabras clave: Data Warehouse, Sistema de Soporte a Decisiones, OLAP, Hefesto, Pentaho.

1. INTRODUCCIÓN

Hoy en día uno de los problemas principales en toda empresa con o sin fines de lucro es la gestión de grandes volúmenes de información, y la forma de explotar dicha información para lograr soporte de las decisiones financieras, administrativas y económicas. Es así que toda empresa actualmente debe mantener un control de la información generada día a día para poder tomar decisiones de una forma óptima.

Particularmente, en las bibliotecas y centros de información surge la necesidad de tener un control sobre los datos generados de diversas fuentes y formatos, de integrarlos para tener una visión general de su situación, y así poder tomar decisiones sobre el manejo de recursos y servicios de bibliotecas (Siguenza-Guzman et al. 2015).

Una solución que permita el análisis, el enlace de diferentes fuentes de datos, la interpretación y la presentación de resultados óptimos y eficientes, es un Data Warehouse (DW). Un DW es definido como “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”, determinadas a partir de los requerimientos del negocio considerando el tiempo (Kimball y Ross, 2002). En efecto, el DW es un software hecho a medida construida a partir de los requerimientos de negocio. Integra todas las fuentes de datos involucradas estratégicamente en la organización con el objetivo de ayudar a una adecuada toma de decisiones (Inmon, 2005).

En la literatura especializada sobre estos temas es posible encontrar diferentes enfoques de usos de un DW para la ayuda a la toma de decisiones en bibliotecas. Por ejemplo, Zucca (2003) implementa un DW para la gestión de la colección de una biblioteca, mientras que Chang y Chen (2006) lo utilizan para analizar y categorizar los requerimientos de los usuarios. Ying et al. (2007) usan un DW para realizar un rastreo Web de las fuentes de datos externas. En un estudio reciente, Luan y Jiang (2014) implementan un DW con el objetivo de ayudar en la gestión de la colección. Todos los autores en general reconocen que los sistemas tradicionales de biblioteca no son suficientes para un adecuado soporte en la toma de decisiones, y que una eficiente gestión de la biblioteca necesita de la creación de sistemas y bases de datos especializados que integren eficientemente toda la información existente en una biblioteca.

El resto del artículo está organizado de la siguiente manera: en la sección 2 se describe la selección y aplicación de una metodología para el diseño de un DW, en la sección 3 se explica detalladamente los pasos para la implementación de la solución y finalmente, la última sección sintetiza las principales conclusiones y trabajos futuros.

2. DISEÑO DE UN DATA WAREHOUSE PARA EL CENTRO DE DOCUMENTACIÓN REGIONAL “JUAN BAUTISTA VÁZQUEZ”

Para implementar un sistema de soporte a las decisiones (DSS por sus siglas en inglés Decision Support System) utilizando técnicas de DW se debe considerar la arquitectura de datos a utilizar, las fuentes de datos disponibles, las herramientas tecnológicas que se utilicen para la integración de los datos y la metodología a seguir para que la construcción de un DW se lleve a cabo exitosamente (Siguenza-Guzman et al. 2014). A continuación se explica la selección de dicha metodología y la aplicación de ésta para diseñar un DSS.

2.1. Selección de la metodología para desarrollar un Data Warehouse

Existen varias metodologías que pueden ser utilizadas para el desarrollo de un Data Warehouse; sin embargo, las que se imponen entre la mayoría son las metodologías de Inmon, Kimball y Hefesto.

Inmon propone una metodología top-down que transfiere la información de los diferentes procesamientos de transacciones en línea (OLTP) a un repositorio centralizado (Inmon, 2005). Kimball por su parte, propone una metodología bottom-up que permite construir un Data Warehouse de forma escalonada considerando los Data Marts ya creados para construir el DW (Kimball y Ross, 2002).

Hefesto es una metodología que combina las mejores prácticas de metodologías existentes como por ejemplo Inmon y Kimball, y permite crear un Data Warehouse a partir de la recolección de requerimientos, seguido de un proceso de extracción, transformación y carga (ETL), hasta definir un esquema lógico sea este un Data Warehouse o Data Mart (Bernabeu, 2010).

Luego de comparar las propuestas de cada metodología, el enfoque seleccionado es el de Hefesto, ya que permite dar énfasis a los requerimientos de los usuarios permitiendo que los resultados persigan los objetivos definidos al momento del planteamiento del Data Warehouse. Hefesto contempla cuatro pasos para la construcción de un DW, como se indica en la Figura 1.

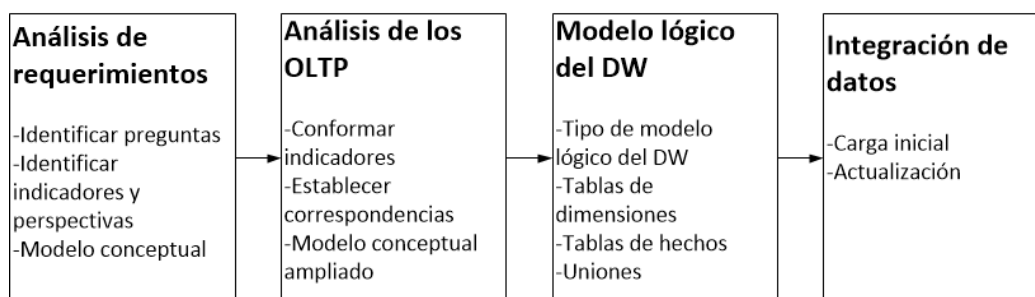


Figura 1: Pasos de la metodología Hefesto

2.2. Aplicación de la metodología Hefesto

La aplicación de la metodología para el diseño del Data Warehouse se realizó en el Centro de Documentación Regional “Juan Bautista Vázquez” (CDRJBV) perteneciente a la Universidad de Cuenca, Ecuador. El Centro de Documentación está conformado por las Bibliotecas de los Campus Central, El Paraíso (Áreas de la Salud) y Yanuncay (Áreas Agropecuarias y de Artes). Su conjunto documental está constituido por publicaciones convencionales en todas sus formas, así como por diversos soportes digitalizados, audiovisuales y bases de datos en línea.

2.2.1. Análisis de requerimientos

En esta etapa se deben conocer los procesos de la organización y recoger los requerimientos de los usuarios. Para el caso de estudio, los datos se recolectaron a partir de los requerimientos internos basados en procesos transaccionales, generados diariamente como se presentan en la Figura 2 así como la matriz de evaluación holística de bibliotecas (Tabla 1) propuesto por Siguenza-Guzman et al. (2015). La adopción de esta metodología de evaluación holística, se debe a que permite un análisis integral tanto interno como externo del centro de documentación, analizando el rendimiento y la calidad de los servicios y la colección de la biblioteca, además de un análisis de uso y comportamiento de los usuarios durante su acceso al material bibliográfico disponible. Para mayor información sobre el procedimiento empleado en la recolección de requerimientos, véase Haro Valle y Pérez Rocano (2014).

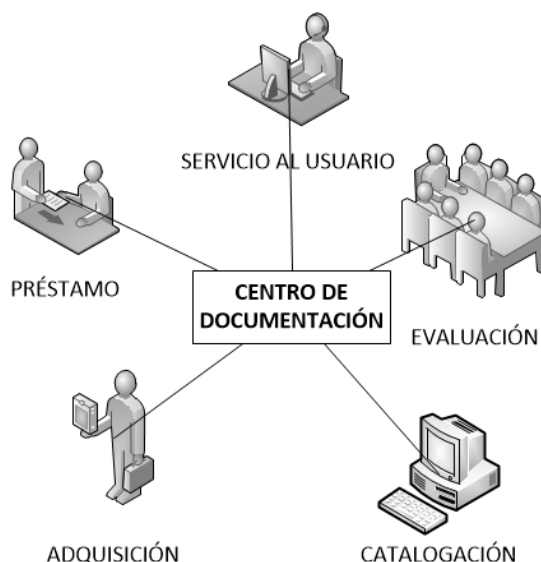


Figura 2: Transacciones diarias del CDRJBV

Tabla 1: Metodologías propuestas para la evaluación de bibliotecas a través de un enfoque holístico (Siguenza-Guzman et al. 2015)

	Sistema Bibliotecario	Uso de la Colección
Perspectiva Interna (Biblioteca)	Análisis de Servicios Procesos, Tiempo, Recursos	Análisis de Uso Información implícita y explícita
Perspectiva Externa (Usuarios)	Análisis de Calidad Datos estadísticos, Buzón de sugerencias, Pruebas de uso, Encuestas de satisfacción.	Análisis de la Colección Patrones de citas, Patrones de publicación, Revistas descargadas, Factor de impacto de revistas

2.2.2. Análisis de las fuentes de datos

Una vez definidos y analizados los requerimientos, el segundo paso es el análisis de los OLTP. Este paso requiere considerar las fuentes de datos tanto internas como externas que permitan establecer correspondencias con los requerimientos. El CDRJBV posee fuentes propias de información que se pueden obtener y explorar sin necesidad de acudir a terceros, ya sean fuentes generadas por su propia

gestión o aquellas fuentes que han sido elaboradas para resolver problemas específicos. Los datos a los que se tiene acceso internamente son los relacionados a los procesos de: Catalogación, Préstamos, Análisis de costos de procesos, Repositorio digital, Acceso remoto (EZProxy), Estadísticas de acceso a bases de datos digitales y Evaluación de servicios. La parte superior de la Figura 3 muestra las fuentes de información internas involucradas en dichos procesos. Además de las fuentes internas, el CDRJBV dispone de acceso a otras fuentes externas provenientes del Departamento de Tecnologías de la Información y Comunicación de la Universidad, relacionadas con ciertos procesos académicos, socioeconómicos y administrativos, como lo indica la parte inferior de la Figura 3.

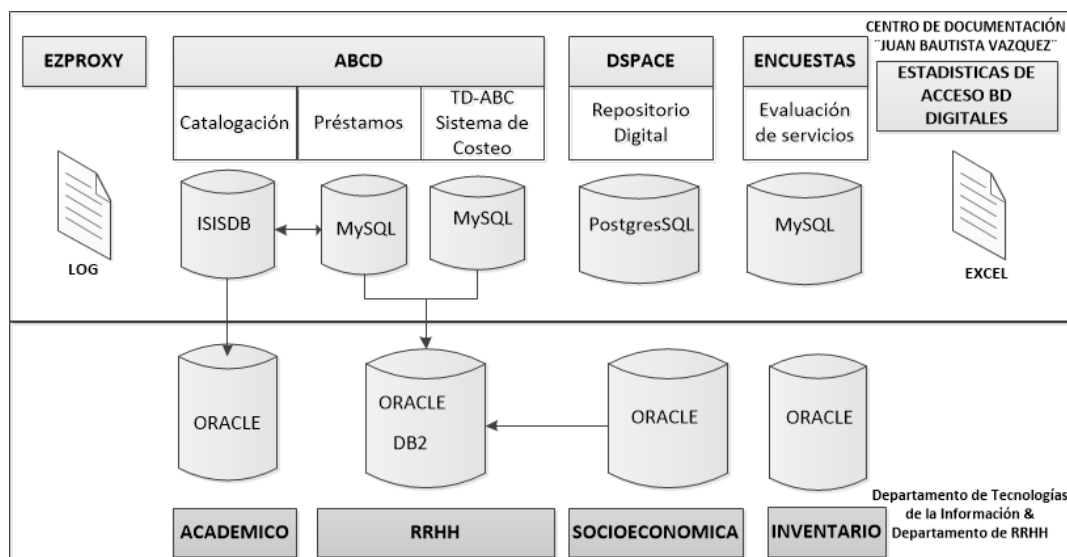


Figura 3: Fuentes de datos CDRJBV.

En este paso, se define la importancia de almacenar los datos que aunque parezcan innecesarios para los procesos transaccionales, serán de suma importancia para generar información que ayude a la toma de decisiones. Al analizar los requerimientos y las fuentes de datos existentes se genera un modelo conceptual. La Figura 4 presenta el modelo conceptual resultante para el CDRJBV.

2.2.3. Modelo lógico del Data Warehouse

Después de analizar los procesos que se llevan a cabo en el CDRJBV y las fuentes de datos existentes, se procede a diseñar el modelo lógico para el Data Warehouse basándose en los cuatro cuadrantes del análisis holístico propuesto por Siguenza-Guzman et al. (2015).

- *Cuadrante I:* En este cuadrante se tiene un total de cinco modelos que representan a cada uno de los principales procesos realizados en el CDRJBV y sus correspondientes costos (ver Figura 4). Los procesos de *Catalogación*, *Reservas de libros*, *Préstamos*, *Préstamos Interbibliotecario* y *Adquisiciones* tienen en común perspectivas de costos, sean estos económicos o materiales. Un ejemplo del modelo lógico del proceso *Préstamos* se presenta en la Figura 5.
- *Cuadrante II:* En este cuadrante se tiene un total de tres modelos que son procesos relacionados a la evaluación de los servicios bibliotecarios (ver Figura 4) entre los que se encuentran: *Atención al Cliente*, *Evaluación* y *Encuestas*.
- *Cuadrante III:* En este cuadrante se encuentra un solo modelo que es el *Repositorio Institucional* (ver Figura 4).

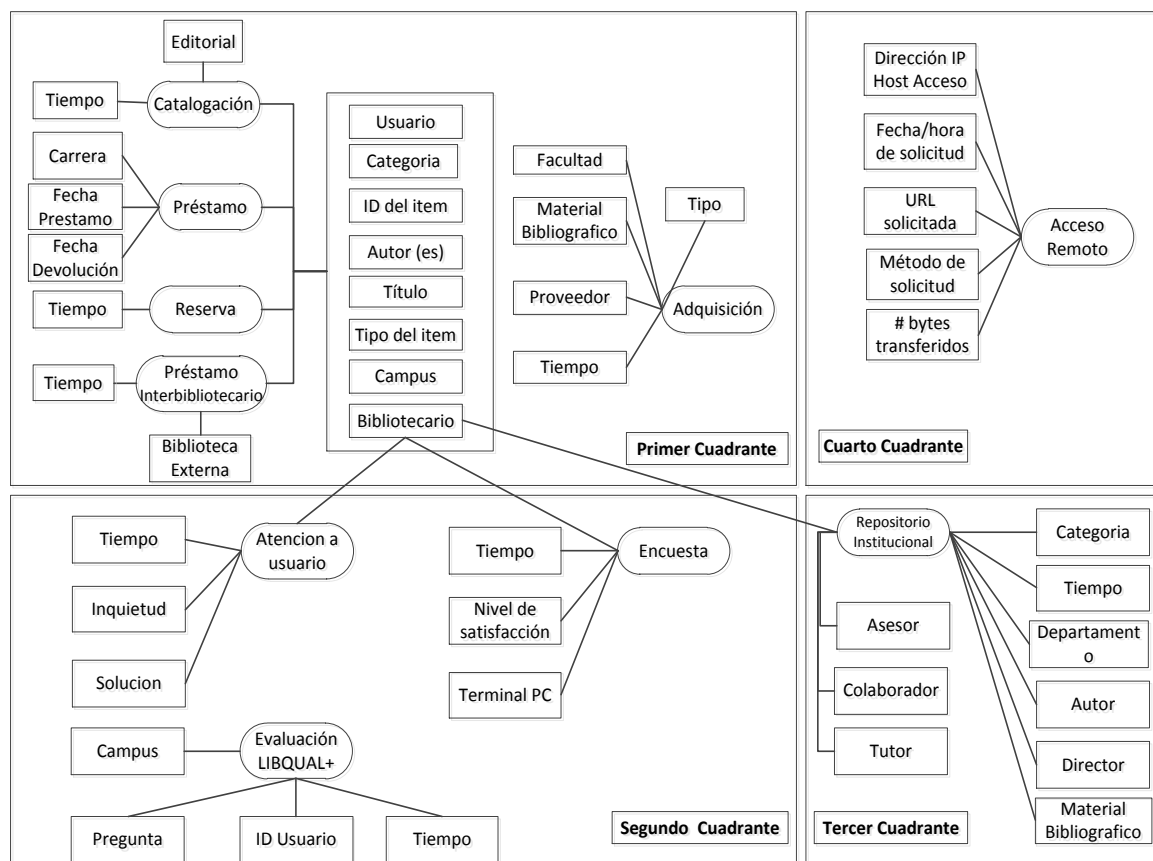


Figura 4: Modelo conceptual del DW para el CDRJBV

- *Cuadrante IV:* Este cuadrante está conformado por el modelo de *Acceso Remoto* a los recursos bibliotecarios (ver Figura 4).

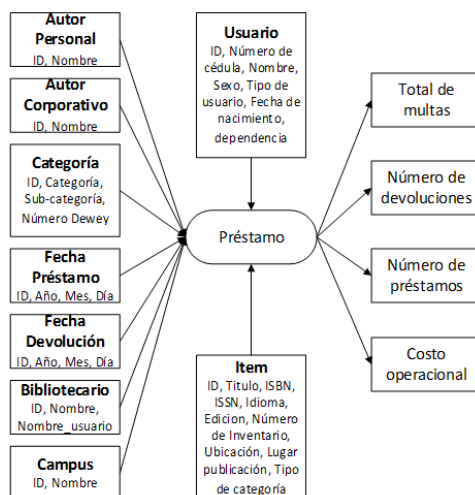


Figura 5: Modelo lógico del proceso Préstamos

2.2.4. Integración de datos

El último paso que propone la metodología Hefesto es la integración de datos en el cual se utilizan técnicas de limpieza y procesos ETL para la carga de datos en el DW, proceso que se detalla en la

sección 3. Después del análisis de requerimientos, éste es el proceso más largo dentro de la implementación del DW.

3. IMPLEMENTACIÓN DE UN DSS PARA EL CENTRO DE DOCUMENTACIÓN REGIONAL “JUAN BAUTISTA VÁZQUEZ”

Una vez diseñado el modelo conceptual y lógico del DW, el siguiente paso es poblar el DW a través de la herramienta correspondiente.

3.1. Selección de la herramienta tecnológica

Siguiendo las políticas gubernamentales en el Ecuador que determinan el uso de software libre en las instituciones del sector público (Ecuador, 2008), en el presente proyecto se consideran las herramientas de código abierto (Open Source) que permiten realizar todo el proceso de creación del Data Warehouse, entre las más destacadas están JasperSoft y Pentaho (Golfarelli, 2009; Dietz y Singh, 2009). En la Tabla 2 se realiza una comparativa más exhaustiva de estas dos herramientas de Inteligencia de Negocios.

Tabla 2: Comparación de Herramientas para la creación del DW

	JasperSoft	Pentaho
Tipo de Licencia	JasperSoft LGPL V3 iReport GPL V3	Pentaho Reporting LGPL V2.1
	Modelo normalizado	
Lenguaje de desarrollo	Java, Perl	Java
Desarrollo de Plug-in	Eclipse Plug-in Available	Eclipse Plug-in Available
JDBC Driver	6 conexiones a fuentes	37 conexiones a fuentes
Compilación de Reporte	SI	NO
Plataformas soportadas	Windows, Linux, Mac OS X	Windows, Linux, Mac OS X
Servidor de Aplicaciones	JBoss	JBoss
Servicios web	Tomcat	Tomcat
Reporte/Gráficos	Si (JasperReport)	Si (Pentaho Report Designer)
Cuadros de Mando	Si (JFreeChart)	Si (JFreeChart)
Herramienta ETL	Si (JasperETL)	Si (Pentaho Data Integration)
Soporte para Minería de Datos	No	Si (Weka)

Jasper maneja un modelo de negocio del tipo comercial de código abierto, ofreciendo soluciones al análisis e integración de datos con una arquitectura flexible construida en un modelo escalable para que sea integrable con otras aplicaciones (Jaspersoft Corporation, 2014). Por otra parte, Pentaho ofrece una suite de software integrado que permite a la empresa desarrollar soluciones orientadas al problema (Dietz y Singh, 2009). Esta herramienta combina muchos componentes que facilitan una adecuada administración del Data Warehouse, entre los cuales están: Mondrian, JFreeReport, Kettle, Weka, entre otros. La herramienta seleccionada para la creación del DW es Pentaho, debido a que tiene una arquitectura que se basa en servidores, motores y componentes, muchos de ellos estándares ofreciendo una plataforma de Inteligencia de Negocios escalable y sofisticada. Pentaho facilita una adecuada creación y administración del DW y permite descubrir nuevo conocimiento a través de su integración con Weka, un completo banco de herramientas para minería de datos, como se indica en la Tabla 2.

3.2. Implementación del componente de integración de información para base de datos documentales

Pentaho posee componentes de conexión a bases de datos relacionales y archivos de diferentes formatos, pero no dispone de una conexión para las bases de datos documentales. Por lo tanto, como parte de la implementación del Data Warehouse se contempló la creación de una herramienta en Java que transforme a registros MARC los datos del material bibliográfico que posee el CDRJBV. MARC4J fue la librería de código abierto seleccionada para la implementación de un proceso que permita generar un archivo en formato MARC21 a partir de la base documental IsisDB.

Al generar el archivo MARC se encontró que la información del material bibliográfico que dispone el CDRJBV tenía una serie de inconsistencias en los registros almacenados. Estas inconsistencias, hicieron que la mayor parte de los datos no fueran válidos. Por ejemplo, se trabajó inicialmente con la muestra que contenía 149.931 registros de material bibliográfico de los cuales 54.755 eran datos válidos; a los registros restantes se dio un tratamiento especial para convertirlos en información útil y así cargar al Data Warehouse. Así mismo, se encontró que varios registros importantes para la creación de la cabecera del formato MARC21 tenían valores vacíos, como por ejemplo, los campos sobre el tipo de registro y el nivel bibliográfico. En la Tabla 3 se resumen el número de registros procesados con los errores encontrados.

Tabla 3: Resumen del número de registros encontrados con errores de acuerdo al campo de consulta

Campo	Nro. de Registros
005	65.138
041	84.626
500	108.127
007	53.770

Una vez que los datos han sido depurados, un archivo de extensión .mrc fue generado con formato MARC21. La herramienta de conversión desarrollada que genera el archivo MARC y que a su vez considera todo el tratamiento de errores se la denomina IsisToMarc-Java.

Finalmente, el archivo .mrc generado es interpretado desde Pentaho Data Integration (KETTLE) leyendo los datos en el formato (MFN-Campo-Subcampo-Dato) como se presenta en la Figura 6 que además indica el proceso realizado.

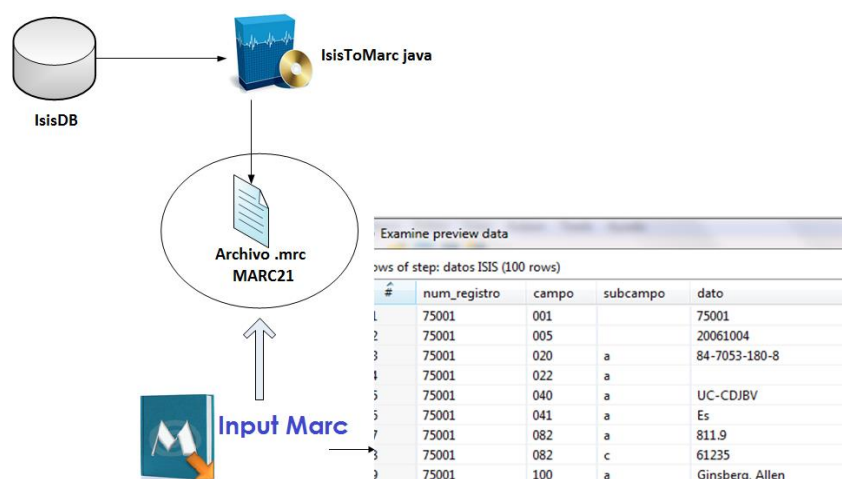


Figura 6: Proceso de conversión de datos almacenados en la base de datos documental IsisDB a un archivo de extensión .mrc

3.3. Procesos ETL

Los procesos ETL permiten extraer los datos a partir de las fuentes de datos OLTP, posteriormente se realiza una limpieza de datos y se transforman para que coincidan con los modelos en los que serán cargados.

Basándose en el diseño del Data Warehouse definido previamente se realizó la integración de la información de los diferentes OLTP's, por ejemplo: los datos de los bibliotecarios y catalogación fueron extraídos desde el repositorio digital y el sistema de gestión de bibliotecas. También se integraron los datos de usuarios y dependencias desde el módulo de préstamos, así como los datos del sistema de adquisiciones, el sistema socioeconómico y el sistema de gestión de personal, los tres últimos pertenecientes a la Universidad de Cuenca. Se aumentó el nivel de detalle de la información en libro, categoría Dewey, dependencias de los usuarios, préstamos interbibliotecarios, reservas, multas, evaluación de bibliotecas y datos de los log's de acceso de los usuarios remotos.

Una vez que se ha realizado la extracción y limpieza de datos se procede a cargar las dimensiones. La Figura 7 presenta un ejemplo del ETL para el proceso de *Préstamos*, donde inicialmente se extraen los registros de la base de datos relacional; para recuperar el *id-libro* se carga los datos del modelo multidimensional de catalogación *dim-libro* previamente almacenada, luego se procede a combinar los registros que se encuentra en el flujo para continuar con las búsqueda de los demás datos y obtener todos los campos necesarios para finalmente depositar en la tabla de hechos.

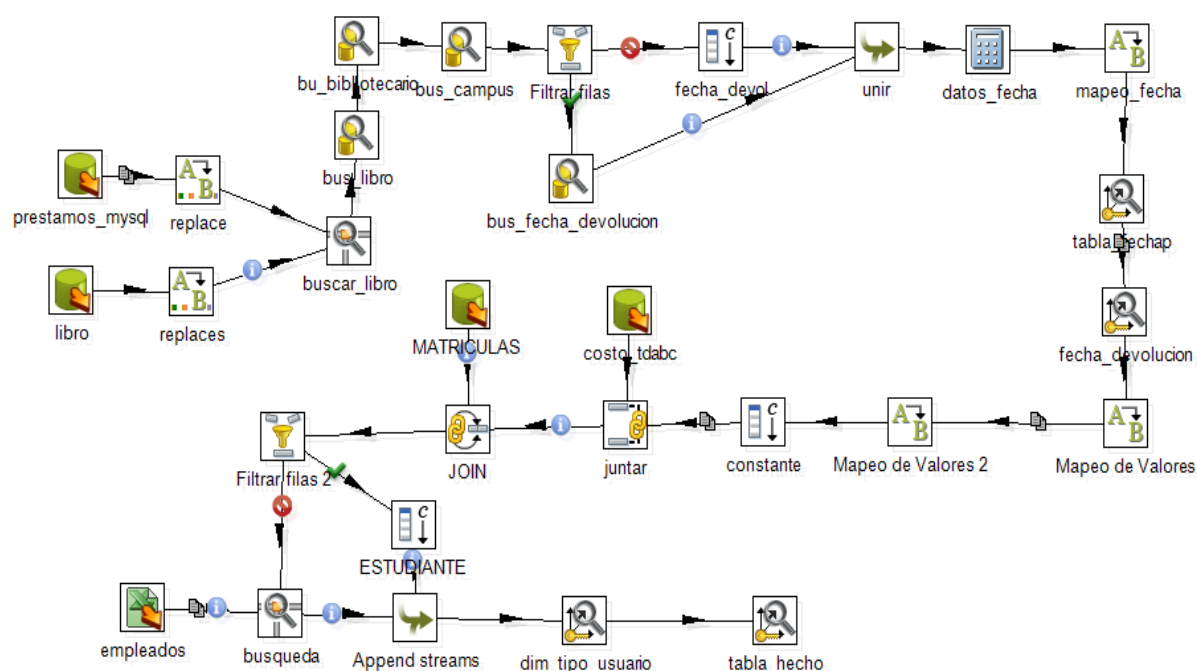


Figura 7: Transformación para el proceso Préstamos

3.4. Generación de cubos

La generación de los cubos se realiza utilizando la herramienta de la suite de Pentaho denominada Mondrian Schema Workbench. Ésta herramienta permite crear esquemas con los cubos deseados, compuestos por dimensiones con sus respectivos atributos almacenados en el modelo multidimensional. Además, se crea la tabla de hechos con los indicadores a los cuales se asigna la función matemática que realizarán. La Figura 8 muestra la estructura del cubo de préstamos generado.

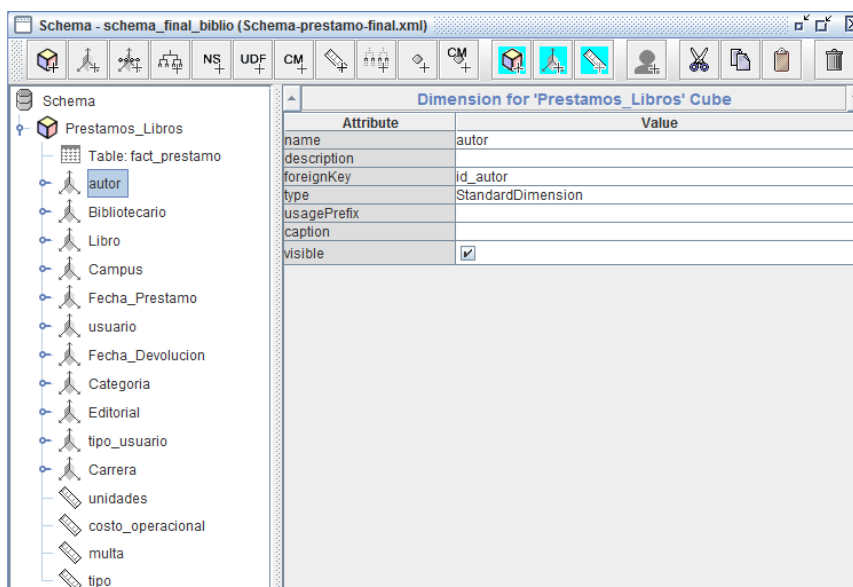


Figura 8: Cubo para el proceso Préstamos

3.5. Pruebas

Después de generar los cubos y publicarlos en el Business Intelligence Server de Pentaho se pueden generar diferentes reportes mediante la selección de las dimensiones deseadas para responder las preguntas planteadas como requisitos, por ejemplo:

¿Cuál es el número de préstamos de un libro de un determinado autor, de una determinada categoría a un tipo de usuario en la biblioteca en una unidad de tiempo?

La Figura 9 muestra un reporte que indica los títulos de los libros disponibles en la biblioteca de los cuales se han solicitado préstamos y se indica la cantidad de préstamos de este libro realizados por un bibliotecario.

Columnas ▼ unidades ⬇		
Filas ▼ titulo ⬇ username ⬇		
Filtro ▼		
Info: 21:29 / 3 x 10577 / 7.37s		
Asia. Gran Malasia. Asia meridional. Asia sudoccidental	retrospectiva	4
	Imontalvan	2
	rmendez	1
	tbermeo	2
	ttorres	1
Asombro	retrospectiva	3
	eduran	1
Aspectos de la biograf?a	rmendez	1

Figura 9: Ejemplo de reporte para el modelo de Préstamos

El ejemplo demuestra que el cubo realizado sobre el modelo multidimensional de préstamos cumple con los requerimientos solicitados, permitiendo generar reportes dinámicos según las

necesidades del CDRJBV. Además, vale recalcar que a pesar de que no se pueda generar consultas sobre dos cubos a la vez, se puede encontrar la información deseada accediendo directamente a los datos del modelo multidimensional mediante consultas SQL. Un ejemplo se presenta en la Tabla 4 que son los datos resultantes al comparar las catalogaciones realizadas por un determinado bibliotecario desde el sistema de gestión de bibliotecas de la colección física (I cuadrante) y en el sistema de repositorio digital (III cuadrante).

Tabla 4: Reporte de libros físicos y digitales catalogados por el usuario

Usuario	Sistema de Gestión de Bibliotecas	Sistema de Repositorio Digital
acriollo	220	0
adelosreyes	30	32823
bcabrera	0	239
dnaula	8	392
gmartinez	250	0
epenafiel	0	92
jmogollon	15	96
lmora	20	576
maguilar	850	0
mfaican	800	0
mbrito	0	10
mgutierrez	9764	357
mquezada	240	0
mmendez	20	924
npena	100	0
rcampoverde	100	0
salvarado	200	0
ttorres	300	0
tbermeo	10	130

4. CONCLUSIONES Y TRABAJOS FUTUROS

Actualmente las bibliotecas disponen de tanta información que parece complicado sacarle provecho a todos sus datos. El Data Warehouse al ser una plataforma que integra la información disponible de una empresa, ayuda como soporte en la toma de decisiones ya que permite tener un panorama global de los procesos. Es por esto que mientras se disponga de más información histórica, los reportes de este nuevo repositorio común serán de mayor ayuda para los departamentos administrativos a los cuales va dirigido el DW.

Adicionalmente, el DW no solo beneficia la integración de información sino que también permite *madurar los procesos* transaccionales que diariamente se llevan a cabo. Por ejemplo, al realizar un análisis de los datos almacenados en las diferentes fuentes de información se descubrió una serie de inconsistencias en la catalogación de los libros, consecuentemente, el uso de estándares y manuales de procesos para el ingreso de información permitirá reducir el tiempo de limpieza de datos, facilitando la población de datos en los modelos multidimensionales.

Entre los desafíos encontrados al momento de implementar el DW se encuentran la poca integración existente entre las distintas bases de datos que poseen el CDRJBV y la Universidad de Cuenca. Por ejemplo, no existen identificadores que logren relacionar el proceso de adquisiciones

llevado a cabo por la Universidad con el proceso de catalogación del Centro de Documentación lo cual imposibilita que se pueda combinar la información para obtener un mayor beneficio.

Otro importante reto que se presenta durante la implementación del DW es el alto costo de tiempo de carga inicial de los datos al ejecutar los procesos ETL debido a la gran cantidad de datos disponibles. Por lo tanto, se recomienda que al momento de desarrollar un DW se considere cargar los datos de acuerdo a un rango de tiempo para tratar de reducir el tiempo de procesamiento y no cargar todos los datos a la vez.

Para incrementar la escalabilidad del DW implementado, el uso de la matriz holística para evaluación de bibliotecas fue muy importante al momento del diseño del DW. Gracias a esto, los procesos de reservas de libros, prestamos interbibliotecarios, manejo de multas y encuestas de satisfacción de usuarios que actualmente no son llevadas a cabo en el CDRJBV pudieron ser considerados para su futura incorporación.

La suite de Pentaho son herramientas necesarias para la implementación de un DW que conjuntamente con la metodología Hefesto proporcionan los pasos necesarios para que la implementación sea eficiente. Sin embargo, hay que considerar que no existe conexión desde esta suite de herramientas a bases de datos documentales como la existente en el CDRJBV y en varias bibliotecas a nivel mundial, por lo que se propone desarrollar plugins para facilitar el acceso a este tipo de bases de datos.

Finalmente, hay que considerar que el beneficio de un DW incrementa cuando se aplica técnicas de minería de datos que ayudan a descubrir conocimiento oculto. Al aplicar éstas técnicas, se puede determinar patrones de comportamiento de los usuarios de los Centros de Documentación o bibliotecas que sería el paso siguiente después de la implementación de un DW.

AGRADECIMIENTOS

Agradecemos a todo el personal del CDRJBV, en especial a la Lic. Rocío Campoverde, al Ing. Mauricio Brito, al Ing. Andrés de los Reyes, y a la Lic. Margarita Gutiérrez, que siempre estuvieron prestos en apoyarnos y asesorarnos en las dudas presentadas. Este trabajo ha sido parcialmente financiado por el Proyecto “Cambio institucional para fortalecer la investigación y la educación” del Programa VLIR-IUC. De igual manera, agradecemos a CEPRA, por el apoyo brindado a través del proyecto “Plataforma de integración, publicación y consulta integrada de Recursos Bibliográficos en la Web semántica”.

REFERENCIAS

- Bernabeu, Ricardo Dario. 2010. Hefesto – Data Warehousing: Investigación y Sistematización de Conceptos - Hefesto: Metodología para la Construcción de un Data Warehouse. Córdova, Argentina. <http://www.businessintelligence.info/docs/hefesto-v2.pdf>.
- Chang, Chan-Chine, y Ruey-Shun Chen. 2006. “Using Data Mining Technology to Solve Classification Problems: A Case Study of Campus Digital Library.” *Electronic Library*, The 24 (3): 307–21. doi:10.1108/02640470610671178.
- Dietz, Benjamin, y Lily Singh. 2009. “Open Source BI Reporting Tool Review.” *Open Source Business Resource*, No. Septiembre 2009. <http://timreview.ca/article/288>.
- Ecuador. 2008. Decreto de La Ley de La República Del Ecuador. Decreto 1014 Para Establecer Como Política Pública La Utilización de Software Libre En El Ecuador. http://www.esepoch.edu.ec/Descargas/programapub/Decreto_1014_software_libre_Ecuador_c2d0b.pdf.

- Golfarelli, Matteo. 2009. "Open Source BI Platforms: A Functional and Architectural Comparison." En *Data Warehousing and Knowledge Discovery*, edited by Torben Bach Pedersen, Mukesh K. Mohania, and A. Min Tjoa, 287–97. *Lecture Notes in Computer Science* 5691. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-03730-6_23.
- Haro Valle, Valeria Alexandra, y Wilson Rodrigo Pérez Rocano. 2014. "Data Warehouse para el Centro de Documentación Regional 'Juan Bautista Vázquez'". Tesis de Pregrado, Ecuador: Universidad de Cuenca. <http://dspace.ucuenca.edu.ec/handle/123456789/19878>.
- Inmon, William H. 2005. *Building the Data Warehouse*. 4 ed. John Wiley & Sons.
- Jaspersoft Corporation. 2014. *Jaspersoft Studio User Guide*. Release 5.6. http://d2553lapexsdlr.cloudfront.net/sites/default/files/docs/jaspersoft-studio-user-guide_0.pdf.
- Kimball, Ralph, y Margy Ross. 2002. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd ed. John Wiley & Sons.
- Luan, Xiu Mei, y He Jiang. 2014. "Design of Books Analysis System in University Library Based on Data Warehouse." *Applied Mechanics and Materials* 513-517 (Febrero): 2121–24. doi:10.4028/www.scientific.net/AMM.513-517.2121.
- Siguenza-Guzman, Lorena, Victor Saquicela, y Dirk Cattrysse. 2014. "Design of an Integrated Decision Support System for Library Holistic Evaluation." In *Proceedings of IATUL Conferences*, 1–12. No. 1. Espoo, Finland.
- Siguenza-Guzman, Lorena, Alexandra Van Den Abbeele, Joos Vandewalle, Henri Verhaaren, y Dirk Cattrysse. 2015. "A Holistic Approach to Supporting Academic Libraries in Resource Allocation Processes." *The Library Quarterly: Information, Community, Policy* (in Press).
- Wah, Teh Ying, Ng Hooi Peng, y Ching Sue Hok. 2007. "Building Data Warehouse." En *Proceedings of the 24th South East Asia Regional Computer Conference*, 15:51–56. No. SP4. Bangkok, Thailand. http://pdf.aminer.org/000/307/364/acquiring_and_integrating_external_data_into_data_warehouses_are_you.pdf.
- Zucca, Joe. 2003. "Traces in the Clickstream: Early Work on a Management Information Repository at the University of Pennsylvania." *Information Technology and Libraries* 22 (4): 175–78.